PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 9 SEPTEMBER 15, 1923 Number 9

A THEORY OF COLOR PRODUCTION.[1] *I*

By Julius Stieglitz

Kent Chemical Laboratory, University of Chicago

Communicated July 5, 1923

For about half a century chemists have sought in the arrangement of atoms in the molecules of organic dyes the source of their intense and beautiful colors. As a matter of fact one can recognize in the structure of the molecule of any dye two fundamental groups of atoms, whose importance was first emphasized by Witt[2]—the *chromophoric* or color-carrying group, and the *auxochromic* or color-intensifying group. In the structure of a simple typical dye, indophenol, given below, the quinoid complex Q is the chromophore, the OH group in the hydroquinoid complex HQ is the auxochrome. Further, basic dyes form salts with acids, phenolic dyes combine with bases and the maximum color depth and intensity are very commonly found in these salts. To illustrate: phenylimido-quinone $O:C_6H_4:NC_6H_5$ is yellow, with the chromophoric quinoid nucleus $:C_6H_4:$; its color is deepened to red by the introduction of an OH group, forming indophenol $O:C_6H_4:NC_6H_4OH$; and with alkali we obtain maximum color intensity and color depth, an indigo blue, through the formation of a salt $O:C_6H_4:NC_6H_4OMe$.

Now, N. R. Campbell,[3] G. N. Lewis,[4] Baly[5] and several others have emphasized the fact that absorption of light waves and production of the complementary color by a dye must be due to vibrations of electrons in the molecules of a dye. But which electrons show just these vibrations, why they vibrate in this manner in the molecules of a dye and not in colorless compounds and what the connection is between these vibrations of electrons and the above established facts connecting structure and color, are questions that have not hitherto been answered, as far as the author is aware.

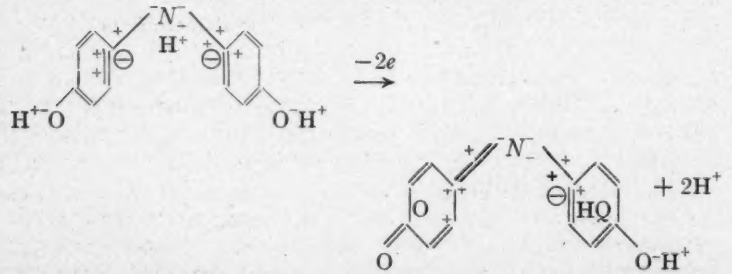
The key to these fundamental questions is found in a fourth fact long known as holding for all dyes without exception: by *reduction* every dye becomes colorless, forming the so-called leuco-dye. By *oxidation* of the leuco-dye, the color is restored. Thus, for indophenol we have:

$$O{:}C_6H_4{:}NC_6H_4OH \underset{-2H}{\overset{+2H}{\rightleftharpoons}} HOC_6H_4.NH.C_6H_4OH$$

colored dye colorless leuco-dye

But oxidation fundamentally is the removal of *electrons from an atom* and by oxidation we in turn invest the oxidized atom with oxidizing power, by means of which it can *attract electrons* and re-absorb the lost electrons, a process we call reduction. A start being made from these facts, which connect the fact of color of a dye with relations involving electrons, an explanation of the color of dyes and of color in general in organic and inorganic compounds has been developed[6] that will be presented briefly here. *It is but logical to correlate the two fundamental relations, oxidation-reduction and color-production because they have in common the fact that they are both of them dependent on the freeing from restraint of valence electrons by positive atoms.*

In the colorless leuco-dye we evidently have a symmetrical molecule containing on *either* side the group $-NHC_6H_4OH$. This is a *strong reducing* group exactly as in aminophenol $H_2NC_6H_4OH$ itself and the equivalent group in the related hydroquinonol HOC_6H_4OH, both of which are used in photography as reducing agents or developers. The leuco-dye is colorless but when we oxidize it, we remove two electrons from a carbon atom[7] at the valence $\ominus$ (as a result the molecule must let go the equivalent $2H^+$ to the oxidizing agent) and the red indophenol dye is produced:[8]



Now, it is clear that we could have oxidized *either* the right *or* the left half of the molecule (either $\pm C\ddagger$ atom linked to $N^{=}$). We have chosen to represent the left side as oxidized, which is thereby converted into an oxidizing complex, the oxidizing power residing in the fully oxidized carbon atom $\ddagger C\ddagger$ of the quinoid complex Q, as indicated in the diagram. But the right half of the molecule with the group NC_6H_4OH could have been oxidized equally well and consequently it must represent an active *reducing group*, still containing the two electrons which might have been removed by oxidation. We find these in the carbon atom $\pm C\ddagger$ indicated in the diagram on the right side of the nitrogen atom, which by the loss of

two electrons would form $\ddagger C \ddagger$. It must be evident, in fact, that the left half (the $\ddagger C \ddagger$ attached to the left of $N^{=}$) of the molecule could oxidize the right half (the $\pm C \ddagger$ attached to $N^{=}$) and we would get the original molecule again, which confirms the existence of oxidizing power in the quinoid group on the left, of reducing power in the hydroquinoid group on the right.

It is the pull of the positive, oxidizing carbon atom $\ddagger C \ddagger$ of the quinoid group Q on the two electrons in $\ominus$ of the reducing carbon atom $\pm C \ddagger$ of the hydroquinoid group HQ, which frees these electrons sufficiently from intra-atomic restraints and makes possible under the excitation of light the longer vibrations which absorb light of visible wave-lengths and produce the complementary color. There is no actual passage of electrons from $\pm C \ddagger$ to $\ddagger C \ddagger$, back and forth—the vibrations are *intra-atomic* and not inter-atomic. The experimental proof for this important conclusion will be given further on. For the present it is sufficient to point out that such an actual passage of electrons to and fro from the $\pm C \ddagger$ to the $\ddagger C \ddagger$ atom as a cause of light absorption would involve *atomic migrations* (H^+ of the OH group in the HQ nucleus to the : O in the Q nucleus) and changes of the other atomic bonds in the nuclei, which would be far too slow to account for the absorption of light waves.

The oxidizing atom $\ddagger C \ddagger$ is found in the quinoid complex, that is the *chromophoric* group of the old structural theory of color. The whole function of the chromophore we must now recognize as forming this positive, oxidizing component in a field of force attracting the electrons of a reducing atom. The reducing atom $\pm C \ddagger$ in turn is found in the hydroquinoid nucleus holding the *auxochromic* group of the old theory. It must be evident now that the auxochrome is introduced into the dye molecule primarily to give this nucleus sufficiently strong[9] *reducing* power: for aniline $H_2NC_6H_5$, whose radicle we find in the yellow O: C_6H_4: NC_6H_5, is not a powerful reducing agent, but when we introduce OH into its molecule we obtain the strong reducing agent aminophenol $H_2NC_6H_4OH$ (a photographic developer). Briefly then, the auxochrome is introduced primarily for the development of strong reducing power in the hydroquinoid nucleus. But *reducing power means freedom for electrons to escape* from an atom under sufficiently strong compulsion and therefore also *freedom to vibrate*, say under the impact of white light, when the forces acting upon them are nicely balanced. We find here this nice balance in the pull on the electrons in $\ominus$ by the positive nucleus of the $\overset{+}{\ominus}C \ddagger$ atom itself and the pull on the same electrons by the positive atom $\ddagger C \ddagger$ in the quinoid complex and *in close proximity* to the atom $\overset{+}{\ominus}C \ddagger$ containing the electrons in question.[10] It is in this nice balance of two neighboring positive fields acting on the valence electrons left in the atom $\overset{+}{\ominus}C \ddagger$ that we have the fundamental explanation of color production. Exactly as

an oxidizing agent, for instance permanganate with its highly positive nuclear atom $Mn\ddagger\ddagger\ddagger^{+}$, can start an actual flow of electrons from a reducing agent in a properly arranged cell, so an oxidizing atom can sufficiently free electrons from their ordinary intra-atomic restraints and permit vibrations which absorb light and produce color. That is the theory briefly expressed and it is believed that ample confirmation of this fundamental conception will be found.

But further, the reducing power of hydroquinol and of organic compounds in general is enormously increased by the addition of a strong base to the reducing agent (we use alkali with hydroquinol as a photographic developer). In an experiment with a chemometer[11] which roughly measures oxidation-reduction potentials, the addition of alkali say to formaldehyde combined in a cell against permanganate increased[12] the reading of the needle from line 3 to line 21 (an artificial scale). Since the chemometer readings represent exponents of activities or of concentration factors, the extraordinary change brought about by alkali may be realized. Addition of a base to indophenol must likewise enormously increase the reducing power of the hydroquinoid nucleus NC_6H_4OH. That means, it must tend to free still more the reducing electrons in $\overset{+}{\ominus}C\ddagger$ from ordinary intra-atomic restraints. As a result, absorption of longer wave-lengths and production of a *deeper* color (indigo blue in place of red) takes place. The electrons in a much larger proportion of the dye molecules must also be activated and color *intensification* takes place. We have in fact with this maximum of the oxidation-reduction potential within the molecule of the dye also maximum color depth and color intensity.

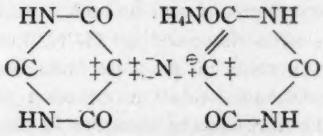
We may summarize our conclusions up to this point briefly as follows: the color of indophenol, a typical organic dye, is due to the combination within a single molecule of a strong positive oxidizing atom $\ddagger C\ddagger$ (in the quinoid or chromophoric group of the structural theory of color) with a strong negative reducing atom $\pm C\ddagger$ (in the hydroquinoid group containing the auxochrome of the old theory) and maximum color depth and intensity is developed by the increase of the reducing power to a maximum with the aid of a strong base. This combination produces color by the absorption of a part of white light through the vibrations of the reducing electrons which are so largely freed from intra-atomic restraints by the factors enumerated.

Let us consider for a moment the further fact that with *basic* dyes, like pararosaniline, we obtain our maximum of color depth and intensity, when the dyes are combined with *acids*. It is a significant fact that the acid is combined with a group in the oxidizing or quinoid complex. Thus, rosaniline $HN:C_6H_4:C(C_6H_4NH_2)_2$ is yellow,[13] but its salt $ClH_2N:C_6H_4:C(C_6H_4NH_2)_2$ is the brilliant red dye of fuchsine or magenta red hue. Now, acids have the same extraordinary effect in intensifying the oxidizing

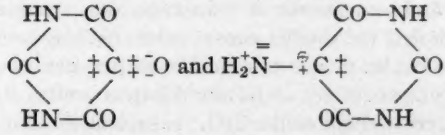
power of very many oxidizing agents that we observed for the effect of bases on reducing power. Thus, dilute sulphuric acid added to the permanganate in the little cell described above raised the needle from line 21 (with alkaline formaldehyde) to line 35! This combination of the acid in salts of basic dyes with the oxidizing or quinoid component of the dye[14] is very characteristic and shows again that we obtain maximum depth of color (acid changed the color from yellow to a brilliant red) and maximum intensity of color when our intra-molecular oxidation-reduction cellule is at its maximum potential. It is further significant that when we combine the reducing hydroquinoid groups $-C_6H_4NH_2$ with acid and thereby destroy their reducing power, we also destroy the brilliancy of color depending on the efficiency of the oxidation-reduction potential (rosaniline trihydrochloride $ClH_2N:C_6H_4:C(C_6H_4NH_3Cl)_2$ is *yellow*).[13]

We must turn now to the evidence that the vibrations of electrons in the dyes are intra-atomic and not due to any transfer to and fro from atom $\pm C \ddagger$ to atom $\ddagger C \ddagger$, from the hydroquinoid nucleus to the quinoid nucleus. This evidence has been brought experimentally by the author with the aid of collaborators[15] in connection with the characteristic purple dye murexide.

In murexide we have the same characteristic combination in close proximity of a reducing atom $\pm C \ddagger$ with an oxidizing atom $\ddagger C \ddagger$ as in the blue indophenol salt. The following diagram brings out this fact:



The dye is indeed prepared by the combination of uramil and its own oxidation product, alloxan, in alkaline solution:



alloxan uramil

Characteristic is that acids break up the dye again into these components. Now, we can prepare a purple murexide by combining dimethyl uramil with alloxan, the group on the right in the diagram containing NCH_3 twice in place of the two NH radicles. This purple dye *maintains its identity*, for it breaks down under the influence of acids into the original dimethyl uramil and alloxan. *Vice versa*, dimethylalloxan combines with uramil in the presence of ammonia to give a different dimethyl murexide (with two NCH_3 groups in place of NH in the *left* ring) which also maintains

its identity and breaks down under the influence of acids into the original components, dimethyl alloxan and uramil.

The source of color production cannot therefore be sought in a migration[16] to and fro of the electrons ⊖ from ‡C‡ to ‡C‡, from one ring to the other, but rather in the *intra-atomic* vibrations of the electrons ⊖ held by the positive core of the ±C‡ atom itself but drawn to the positive oxidizing atom ‡C‡ so close by! Under the excitation of white light the electrons vibrate, absorb part of the light and the deep purple-red color results.

Confirmation of these views developed thus first for the organic dyes is found in the inorganic field and makes the theory one of broadest application.

NOTE: All references will be found at the end of the second part of this report.

A THEORY OF COLOR PRODUCTION. II, INORGANIC COMPOUNDS

BY JULIUS STIEGLITZ

KENT CHEMICAL LABORATORY, UNIVERSITY OF CHICAGO

Communicated July 5, 1923

Aside from colors produced by loosely held electrons in atoms with an odd number of valence electrons and in such atoms as those of metallic elements, which have been described by G. N. Lewis,[4] we may ascribe color in inorganic compounds in part to *intra-atomic* and in part to *inter*-atomic (*intra-molecular*) oxidation-reduction potentials of sufficient force to make possible vibrations of electrons in the reducing component long enough to absorb visible light waves and show their complementary colors. We also have instances where color is produced simultaneously by *intra*-atomic and *inter*-atomic or *intra*-molecular potentials.

Let us consider first the simpler case of color resulting from inter-atomic forces: Salts and oxides of ions of elements with weakest *oxidizing* power of the positive component are uniformly colorless (unless a colored cation like Cr^{+++}, or a colored anion like $CrO_4^{=}$ is present). Thus the salts and oxides of the alkali metals, of the alkaline earth metals, of aluminum are all colorless (with the exceptions noted); their positive ions, K^{+}, Ca^{++}, Al^{+++}, etc., also have the smallest tendency to revert to the metallic condition, i.e., they have the weakest electron-absorbing or oxidizing tendencies. Similarly the positive nuclear atoms in the anions $P‡‡ + O_4^{=}$ and $S‡‡‡O_4^{=}$ have very weak oxidizing power and $PO_4^{=}$ and $SO_4^{=}$ do not produce color as a rule. Again, fluorides are colorless unless combined with a color producing cation, like Cr^{+++}; the fluoride ion holds its electron with tremendous tenacity, and has practically no reducing power of

moment. On the other hand, combination of a strong oxidizing cation with a strong reducing anion leads to strong color development, as in the black platinic iodide $Pt^{++++}I_4^-$. Must we not ascribe the power to absorb light and produce color to the vibrations of one or more valence electrons of I^-, *tending* to reduce the Pt^{++++}? (But there is no actual reduction under ordinary conditions—which confirms the conclusion reached above that color does not depend on oscillations of electrons to and fro from atom to atom but on intra-atomic vibrations made possible by a neighboring oxidizing positive atom.) Again, in series like Ag^+Cl^- (white), Ag^+Br^- (pale yellow), Ag^+I^- (bright yellow) and $Zn^{++}O^{--}$ (white), $Cd^{++}O^{--}$ (brown), $Hg^{++}O^{--}$ (red), and $Zn^{++}S^{--}$ (white), $Cd^{++}S^{--}$ (yellow) and $Hg^{++}S^{--}$ (red or black), color depth very evidently goes parallel with inter-atomic oxidation-reduction potentials. In all of these series the cations have lost all of their valence electrons and there can be no complicating factors of intra-atomic source as we shall find in an atom like $Cr^{+++}_{\cdots}$. The anions in the series have their full complement of valence electrons and are therefore not in a condition of partial oxidation or reduction.

That the electrons in the anions of these colored compounds are loosely held and have a tendency to reduce the cations is shown by the sensitiveness of the silver halides to light, and by the decomposition of mercuric oxide by heat into Hg° and O°. Heat with its close relations to the infrared wave-lengths of light must increase the tendency of electrons toward the point of passing over to the cation, and as a result we find the white $Zn^{++}O^=$ *turning yellow when heated*—an instance of a common fact.

Compounds containing atoms which have lost *all* of their valence electrons when colored must owe their color to this type of inter-atomic oxidation-reduction potential. Thus, in permanganates $K^{+-}O^{-+}Mn\ddagger\ddagger\ddagger^{=}O_3$, in chromates $(K^{+-}O^{-})_2\ddagger Cr\ddagger\ddagger^{=}O_2$ we find no other source of color than in the tendency of the valence electrons of the oxide atoms $O^=$ to pass to the strongly oxidizing atoms $Mn\ddagger\ddagger\ddagger^+$ and $Cr\ddagger\ddagger\ddagger$. In fact, when these salts are heated, oxygen does pass off as in the case of $Hg^{++}O^=$. But why are perchlorates $K^{+-}O^{-+}Cl\ddagger\ddagger\ddagger^{=}O_3$, sulphates $(K^{+-}O^{-})_2\ddagger S\ddagger\ddagger^{=}O_2$ colorless? The chemometer gives the answer: Equivalent solutions of permanganate and perchlorate against the same alkaline formaldehyde solution gave readings of line 18 for the perchlorate and of 38 for the permanganate; addition of equivalent amounts of dilute sulphuric acid to the two gave line 30 for the perchlorate, but line 62(!) for the permanganate. In other words, as has been already emphasized, quantitative factors control the question of visible color production, and not only the qualitative complex.

Inorganic compounds containing a given atom of an element in two conditions of oxidation frequently give us the most intense colors:[17] ferro-

ferric oxide is black (magnetite), ferro-ferricyanide is deep blue (prussian blue), plumbous plumbate is bright red (minium). That there is a tendency for the reduction say of ferri-atoms by ferro-atoms can be demonstrated by the strong potential of a cell containing a ferrous salt against a ferric salt, but that there is vibration without actual transfer to and fro, as some have assumed,[17] is evident from the behavior of the intensely colored compounds enumerated above (mercuric oxide, permanganate, etc.).

All of the above cases represent color production resulting from *inter-atomic* oxidation-reduction potentials. Another very important group of colored substances owe their color to *intra-atomic* forces. The case of the intensely colored chromic salts will sufficiently illustrate this group. Cr^{+++} has practically no oxidizing power. But in the Cr^{+++} atom *there are still three valence electrons* (indicated in the symbol $\overset{\cdot}{\underset{\cdot}{\overset{+}{\underset{+}{Cr}}}}$). We have here an intra-atomic oxidation-reduction potential evidently of sufficient moment to make possible vibrations of the three valence electrons under the influence of light, which absorb part of the light and produce the complementary colors.

In support of this interpretation of the source of color of chromic salts we have the fact that Cr^{+++} and Al^{+++} salts show an extraordinary chemical and physical resemblance to each other (for instance in the alums) except in the matter of color; and the difference between the two is that the aluminum atom has lost all of its valence electrons in Al^{+++}, while the chromium atom still has half of its valence electrons in $\overset{+}{\underset{+}{Cr}}$. That it is not a question of atomic mass and the intra-atomic relations that go with it, is shown by the fact that salts of the gallium-ion Ga^{+++} are colorless like those of aluminum (Ga is 70; Cr is 52).

It is quite consistent with the tenets of the electron theory of valence that *inter*-atomic forces, such as combination with other atoms, should influence the *intra*-atomic arrangement of the three positive charges and the mobility of the three electrons remaining in the chromium atom $\overset{+}{\underset{+}{Cr}}$. Two arrangements of major stability would thus account for the two series of chromic salts, the green and the violet.

Not all atoms which retain a part of their valence electrons develop color: thus $\overset{+}{\underset{+}{P}}$ in phosphites, $\ddagger\ddagger S$ in sulphites, etc., are not colored. It is chiefly the atoms of metals which show this source of color development and as is so well-known, the intra-atomic restraints on the valence electrons are weakest in the metals.

Obviously, we may have color production simultaneously by inter-atomic causes and by intra-atomic sources of the type just now depicted. The deep green manganates Me_2MnO_4 probably represent an instance of this kind: an easily lost valence electron still remains in Mn^{+++} and

at the same time $Mn\ddagger\ddagger\ddagger O^{-}_{4}$ is a powerful oxidizing agent easily liberating some of its oxygen (O^{-}) as free oxygen (O°).

It is proposed to make an exhaustive study of colored compounds, organic and inorganic, in order to group them as far as possible in the four classes which have been outlined: the Lewis group of atoms with loose odd electrons; the group whose color is due primarily to *intra*-atomic oxidation reduction potentials (as $\overset{+}{\underset{+}{\ddagger}}Cr\overset{\bullet}{\underset{\bullet}{}}$); those whose color is the result of *inter*-atomic potentials (as $Hg^{++}O^{-}$), and those whose color is in part due to inter-atomic oxidation-reduction tendencies, in part due to the intra-atomic type resulting from the presence in a partially oxidized atom of some of its valence electrons.

NOTE: I am glad to learn from my colleague, Professor W. D. Harkins, that the theory as presented would lend itself logically to the Bohr theory of the absorption of light by atoms. For example, if the effect of the attraction of the positive oxidizing atom, say $\ddagger C\ddagger$, on the electron or electrons of the reducing atom, say $\overset{+}{\ominus}C\ddagger$, is to raise these electrons to a higher orbit, each electron in question would be less tightly bound to the nucleus of its atom than would otherwise be the case (this would be the equivalent of the "freeing from intra-atomic restraints" used in the article). The result of this would be that in the absorption of light less energy would be required to lift the electron to a still higher orbit with the consequence that the wave-length absorbed would be of lower frequency. This would shift the absorption from the ultra-violet to the visible spectrum and toward the red end of the latter, and the complementary reflected or transmitted light toward the violet end—a result in harmony with the theory and the experimental work supporting it. A further development of these relations from the point of view of atomistics and as far as possible from a quantitative standpoint will be attempted by us in collaboration. J. S.

[1] The theory was first presented in public in an address before the Minnesota Section of the American Chemical Society in December 1920 (cf. *J. Amer. Chem. Soc.*, Proceedings, **43**, 36) and formed the central feature of the Willard Gibbs Medal Address before the Chicago Section of the society, May 25, 1923. It will be published in full in a series of articles in the *Journal* of the Society, to which reference is made for critical discussions of the work of other authors and of assumptions and conclusions briefly outlined here.

[2] Otto N. Witt, *Ber. D. chem. Ges.*, **9**, 522 (1876).

[3] N. R. Campbell, *Modern Electrical Theory*, pp. 315–6 (1913).

[4] G. N. Lewis, *J. Amer. Chem. Soc.*, **38**, 783 (1916) and these PROCEEDINGS, **2**, 586 (1916).

[5] Baly, *Trans. Chem. Soc.* (*London*), **107**, 248 (1915).

[6] It is not assumed that the theory excludes other sources of color through electron vibrations but it is thought to apply to an extraordinarily large number of cases of organic and inorganic compounds and that the remaining instances will be found to be variations of the fundamental conception in its broadest aspects.

[7] These questions of detail are substantiated at length in the articles which will appear in *J. Amer. Chem. Soc.*

[8] Only the essential valence charges are indicated. The electrons which may escape by oxidation are indicated by the symbol $\ominus$. We recall that 2 electrons are lost when a negative charge is converted into a positive charge.

[9] Vibrations which can absorb light waves of definite length involve *quantitative* limits in the field of forces acting upon the electrons. In the absence of quantitative measurements of intramolecular potentials, the quantitative factors are at least indicated by the terms used. The range of wave-lengths for visible color is extraordinarily narrow, as is well known.

[10] The negative charge on the nitrogen atom $\dot{-}N^{-}$ which holds the oxidized atom $\ddagger C\ddagger$ and the reduced atom $\ddagger C \dot{-}$ close together presumably repels the electrons in $\ddagger C \dot{-}$ and its force is added to the intra-atomic pull of the positive nucleus of the $\ddagger C \dot{-}$ atom itself. But it is thought that the effect of $\dot{-}N^{-}$ is of minor moment because it is an ammono-nitrogen atom as in ammonia with practically *no reducing power* of moment—as little indeed as the reducing power of $O^{=}$ in water—presumably on account of the stability of the nitrogen atom with its full complement of eight electrons found in $N^{=}$. The effect of oxidizing-reduction potentials depends on the *activity* (perhaps the mobility or freedom) of charges according to all the evidence we have.

[11] W. Ostwald, *Zs. phys. Chem.*, **15**, 399 (1894). A description is given in Stieglitz *Qualitative Analysis*, Vol. I, p. 253 (1911) (Century Co.).

[12] The explanation of this effect is given in the author's *Qualitative Analysis*, Vol. I, p. 292 (1911).

[13] Nietzki, *Chemie der Organischen Farbstoffe*, p. 162 (1906).

[14] A combination of hydroquinol against quinone, both in sodium nitrate solutions, showed a chemometer reading of 5; addition of sodium hydroxide to the hydroquinol raised the needle to line 13, and addition of acid to the quinone lifted the needle to line 17. The experiment has only qualitative significance, but clearly supports the views expressed.

[15] Stieglitz and Slimmer, *Amer. Chem. J.*, **36**, 661 (1904) and a forthcoming article by H. V. Tartar and Stieglitz in the *J. Amer. Chem. Soc.*

[16] Certain authors have assumed that there is inter-atomic migration of electrons in color production. Cf. Adams and Rosenstein, *J. Amer. Chem. Soc.*, **36**, 1452 (1914) and H. L. Wells (see the next reference).

[17] H. L. Wells, *Amer. J. Sci.*, **3**, 417 (1922) gives a number of excellent illustrations of this relation. Cf. Willstaetter and Piccard, *Ber. D. chem. Ges.*, **41**, 1465 (1908).

THE ASYMMETRY IN THE DISTRIBUTION OF STELLAR VELOCITIES

By Gustaf Strömberg

Mount Wilson Observatory, Carnegie Institution of Washington

Communicated July 17, 1923

The outstanding features in our knowledge of stellar velocities are the existence of stream-motion, the dependence of the velocities upon absolute magnitude and spectral type, and the asymmetry in the velocity-distribution. The asymmetry in the distribution of velocities was first found by B. Boss[1] in a study of the stars of measured parallaxes and radial velocities. It was independently found by Adams and Joy[2] in studying the space-velocities of 37 stars of large radial velocities. The general

character of this phenomenon was established by the present author[3] from a study of the distribution of the three-dimensional velocity-vectors of 1300 stars of spectral types F to M. The same general effect was found by Boss, Raymond, and Wilson[4] from a study of the space-velocities of 520 stars of measured parallaxes.

This asymmetry is best seen among stars of very large velocity and shows itself as a tendency for stars of high speed to move towards one hemisphere of the sky, namely, that limited by galactic longitudes 160° and 340°. Its effect can be traced even among stars of ordinary speed and produces a difference between the most frequent velocity-vector and the mean velocity.[3] It was suggested by Oort[5] that the asymmetry might be due to the existence of two groups of stars of different internal velocities and of different group-motion. This division into two groups is not distinct and the phenomenon seems rather to be a continuous change in the group-motion with increasing internal velocity.

The purpose of this communication is to show that the asymmetry is a general phenomenon among all the objects in the universe for which we have data regarding velocities, and that it may be reconciled with the existence of a fundamental system of reference for the velocity-vectors. This was found by a deductive course of reasoning, although it might just as well have been found inductively.

Suppose for the moment that there exists a fundamental system of reference, such that excessive velocities referred to it are very infrequent. To express this in a mathematical form we may assume that the velocities referred to this system are distributed according to a spherical distribution-law. The probability of the occurrence, of a velocity (ξ, η, ζ) referred to this fundamental system of reference, is thus proportional to

$$F_1 = e^{-k^2(\xi^2 + \eta^2 + \zeta^2)} \qquad (1)$$

As a coördinate system we shall use the galactic system, the X-axis towards 0° galactic longitude, and the Z-axis towards the north galactic pole. All the velocities have been corrected for a standard solar velocity of 20 km./sec. towards the point $\alpha = 270°$, $\delta = +30°$. The sun's velocity-components in this system of coördinates[3] are then $x = +17.0$, $y = +7.4$, $z = +7.4$ km. Referred to this standard origin, we can write equation (1)

$$F_1 = e^{-k^2[(x + x_0)^2 + (y + y_0)^2 + (z + z_0)^2]} \qquad (2)$$

where x_0, y_0 and z_0 are the velocity components of our origin referred to the fundamental system of reference.

In the absence of this external effect the velocities of all groups of stars are supposed to be distributed symmetrically around a common origin, whose coördinates in our adopted system are α, β and γ. In making this assumption we neglect the difference in group-motion between the central

group and the Taurus group,[3,6] which difference for our purpose is of little importance. As a general expression for a symmetrical velocity-distribution we shall use the function

$$F_2 = \frac{1}{\pi^{3/2}} \sum_v N_v l_v m_v n_v e^{-l_v^2(x'-\alpha')^2 - m_v^2(y'-\beta')^2 - n_v^2(z'-\gamma')^2} \qquad (3)$$

This function expresses the sum of a series of concentric and co-axial ellipsoidal distributions. The velocity α', β', γ' is the velocity of the common centroid, referred to the principal axes and to our previously adopted origin. This division into several ellipsoidal distributions is not necessarily a physical one, but may be regarded as a purely statistical representation, although we may regard each one as a group, which is in a certain sense homogeneous.

Referring our velocities throughout to the common principal axes, we shall for the present drop the accents.

If we now suppose that the frequencies of the actual velocities depend simultaneously upon the two probabilities F_1 and F_2, we find the probability of a certain velocity (x, y, z) expressed by the product of both,

$$F dx dy dz = e^{-k^2[(x+x_0)^2 + (y+y_0)^2 + (z+z_0)^2]} dx dy dz \times \sum_v A_v e^{-l_v^2(x-\alpha)^2 - m_v^2(y-\beta)^2 - n_v^2(z-\gamma)^2} \qquad (4)$$

The function F can be written

$$\left.\begin{aligned} F &= \sum_v B_v e^{-h_1^2(x-c_1)^2 - h_2^2(y-c_2)^2 - h_3^2(z-c_3)^2} \\ h_1^2 &= l_v^2 + k^2 \qquad c_1 = -2a_v^2 k^2 (x_0 + \alpha) + \alpha \qquad a_v^2 = \frac{1}{2h_1^2} \\ h_2^2 &= m_v^2 + k^2 \qquad c_2 = -2b_v^2 k^2 (y_0 + \beta) + \beta \qquad b_v^2 = \frac{1}{2h_2^2} \\ h_3^2 &= n_v^2 + k^2 \qquad c_3 = -2c_v^2 k^2 (z_0 + \gamma) + \gamma \qquad c_v^2 = \frac{1}{2h_3^2} \end{aligned}\right\} \qquad (5)$$

This function represents a sum of ellipsoidal distribution-functions, whose centers are shifted in a direction nearly opposite to that of the "absolute" translation of our origin, by an amount which increases with the internal motion of the group. For stars of small relative motions, for which l, m and n are large as compared with k, the center of the frequency function approaches the limit

$$x = \alpha, \quad y = \beta, \quad z = \gamma$$

For stars whose relative velocities are extremely large the center of the frequency-function approaches the other limit

$$x = -x_0, \quad y = -y_0, \quad z = -z_0$$

In a general way this displacement of the center of the velocity-distribution with increasing internal motions is borne out strongly by stars, globular clusters and spiral nebulae. The velocity-distribution for stars of spectral type F to M for which space-velocities have been computed from proper motions, radial velocities and spectroscopic parallaxes is represented by the sum of two ellipsoidal frequency-functions and the method derived in a previous study[3,6] was employed. The only difference is that the equations of condition used in these investigations were given additional weights inversely proportional to the square roots of the numbers of velocity-vectors terminating in the different parallelopipeds. Probably there are many more than two homogeneous velocity-groups, and the splitting up in two groups may be entirely artificial, but a representation by the sum of two ellipsoidal frequency-functions represents the data sufficiently well. There exists a group of stars of space-velocities exceeding 150 km./sec., which seems to form a separate group, and they were treated accordingly as group III in the table. Groups I and II are thus stars of moderate speed. Group IV consists of the globular clusters and group V of the spiral nebulae. For the two latter groups the radial velocities alone are available and the distribution is supposed to be spherical. Lundmark has computed for other purposes these two groups from data compiled by him, mainly from Slipher's determination, and has kindly allowed me to use his computations.

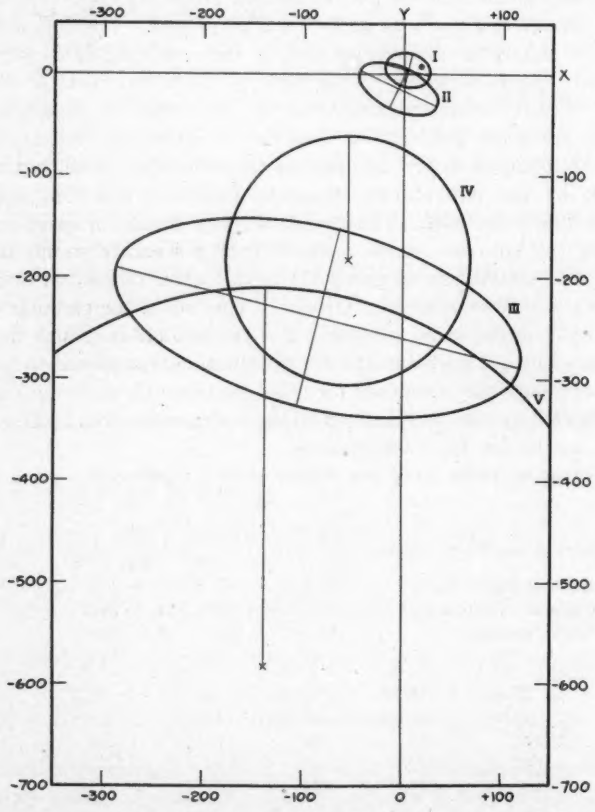
The following table gives the results of this analysis.

		No.	c_1	c_2	c_3
I	Stars of moderate velocity	768	+ 3.6	+ 4.4	+ 0.8
II		510	− 5.3	− 14.8	− 0.4
III	Stars of high velocity.....	21	− 85	−240	− 10
IV	Globular clusters.....	17	− 55	−184	− 39
V	Spiral nebulae.....	42	−137	−583	−220

	a	b	c	L_1	B_1	B_2
I.....	22.66	16.56	10.52	165.4°	−0.3°	+ 2.7°
II.....	42.09	22.10	27.84	155.3	+1.1	−12.7
III.....	210	84	89	167	+5	+ 3
IV.....	122	122	122			
V.....	373	373	373			

The centers of the velocity groups are given in km./sec. referred to the standard origin and to the axes defined in the beginning. The dispersions a, b, and c (mean square velocities) along the principal axes are likewise expressed in km./sec. L_1 and B_1 are the galactic longitudes and latitudes of the major axis a, and B_2 is the galactic latitude of the axis, along which the dispersion is equal to b. In the diagram are plotted the intersections of the velocity-ellipsoids with the galactic plane for the five different groups. The velocity of the sun is denoted by the symbol ⊙. The general tendency for the center of the velocity-distribution to be shifted in the general di-

rection of galactic longitude about 250°, by an amount which increases with the internal velocities of the objects in the group, is clearly shown. According to equation (5) this shift ought to be a linear function of b^2, as the axis b is nearly parallel to the shift itself. Although this is not strictly the case, there can be no doubt about the general tendency. Fur-



Intersections with the galactic plane of the velocity-ellipsoids for different groups of stars: I and II, stars of moderate velocity; III, stars of high velocity; IV, globular clusters; V, spiral nebulae. The diagram shows the progressive change in group-motion with increase in the internal motions of the respective groups. Ordinates and abscissae indicate kilometers per second.

thermore the quantity k may not be strictly the same for the different groups, even if x_0, y_0 and z_0 are constant.

This phenomenon cannot be due to a local gravitational field, as it holds even in the most remote regions accessible to observation. It

seems rather to be a property of something in space or of space itself, a property which formally at least can be expressed as a velocity-restriction in a certain frame of reference. We cannot determine the velocity of our local system in this fundamental frame, since we do not know the quantity k. If the spiral nebulae, however, are supposed to have reached the ultimate velocity which a celestial object can attain, provided such a limit exists, the sun's velocity relative to the spiral nebulae represents at the same time its velocity in this fundamental frame. This velocity of the sun expressed in equatorial coördinates is

$$\alpha = 305°, \quad \delta = +75°, \quad V = 651 \text{ km./sec.}$$

with an uncertainty in the position of the apex of about 15° and in the speed of about 100 km.

Although the phenomenon studied does not prove the existence of a fundamental frame of reference, it is clear that it can be explained as an effect of the existence of such a frame. Other explanations may very well be found. At all events it seems to be a very fundamental phenomenon in which all celestial objects participate, and equally well pronounced in the neighborhood of the sun and in the regions of space occupied by the star-clusters and the spiral nebulae.

[1] *Pop. Astr., Northfield, Minn.*, **26**, 1918 (686); *Pub. Amer. Astr. Soc.*, 22nd Meeting, 1918.

[2] *Mt. Wilson Contr.*, No. 163; *Astroph. J., Chicago, Ill.*, **49**, 1919 (179).

[3] *Mt. Wilson Contr.*, No. 245; *Astroph. J., Chicago, Ill.*, **56**, 1922 (265).
Mt. Wilson Comm., No. 79; these PROCEEDINGS, **8**, 1922 (141).

[4] *Astr. J., Albany, N. Y.*, **35**, 26 (No. 820), 1923.

[5] *Bull. Astron. Inst. of the Netherlands, Leiden*, No. 23, 1922.

[6] *Mt. Wilson Contr.*, No. 257; *Astroph. J., Chicago, Ill.*, **57**, 1923 (77).

ON THE MEAN ABSOLUTE MAGNITUDES OF THE K AND M GIANTS AND THE SYSTEMATIC ERRORS IN TRIGONOMETRIC PARALLAXES

BY WILLEM J. LUYTEN

HARVARD COLLEGE OBSERVATORY, CAMBRIDGE, MASSACHUSETTS

Communicated July 10, 1923

In order to obtain a reliable value for the mean absolute magnitudes of the K and M giant stars we must first establish a system of parallaxes sensibly free from systematic errors. These systematic errors cannot be determined directly by comparing the observed values of the parallaxes among themselves; they are generally found by comparing the observed

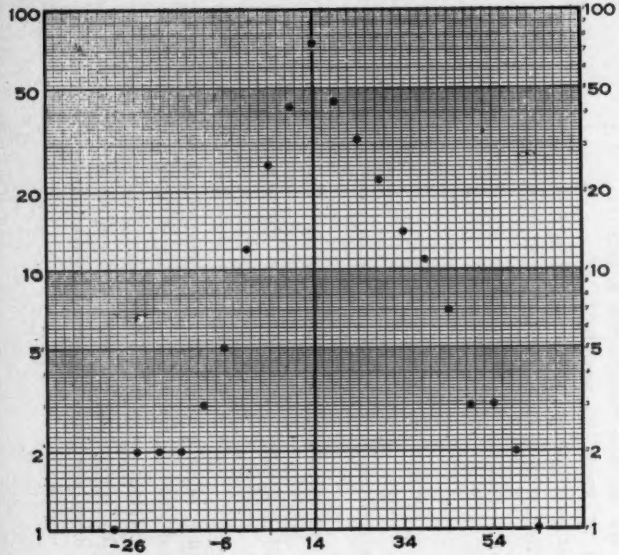


FIGURE 1

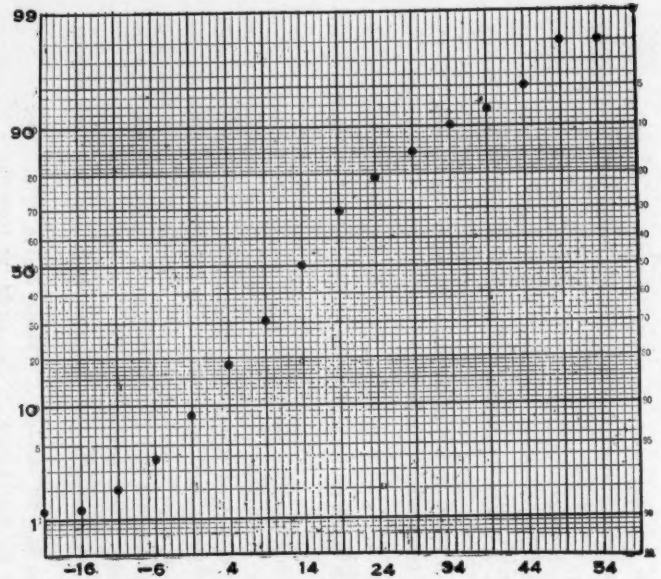


FIGURE 3

FIGURE 2

Figure 1. Distribution of reduced parallaxes larger than median value, on arithmetic-logarithmetic scale. Abscissae: Reduced parallaxes. Ordinates: Number of stars with parallaxes *smaller* than abscissa.

Figure 2. Distribution of reduced parallaxes smaller than median value, on arithmetic-logarithmetic scale. Abscissae: Reduced parallaxes. Ordinates: Number of stars with parallaxes *larger* than abscissa.

Figure 3. Distribution of reduced parallaxes on probability scale. Abscissae: Reduced parallaxes. Ordinates: Percentages of stars with parallaxes smaller than abscissa.

parallaxes with values derived either from parallactic motion or from a comparison of cross motions and radial velocities.

In a recently published memoir,[1] van Rhijn has given an exhaustive investigation of the relation between mean parallax, proper motion, and apparent magnitude, for stars of different galactic latitudes and different spectral classes. He reached the following conclusion, which is opposed to ideas at present generally accepted:

"Although the parallaxes of many stars have been recently determined, the proper motion of which is smaller than $0''.200$, the results derived from these observations can be shown to be systematically too large."

The mean parallaxes of stars with proper motions smaller than $0''.050$ per annum were derived as follows by van Rhijn:

1. From a comparison of the distribution of the proper motions and the linear motions perpendicular to the line of sight.
2. From the mean parallactic motion.
3. From a comparison of the mean radial velocity, freed from the solar motion, and the mean value of the τ component of the proper motion.

It thus seems practically impossible to question the reality of the discordances found by van Rhijn. On account of their fundamental importance in stellar astronomy, however, it may be useful to investigate the matter from another point of view.

The fundamental assumptions underlying van Rhijn's investigation are that the distributions of the radial velocities and of the components of the proper motions are comparable with regard to direction and size. Instead of these assumptions we can make others, no less plausible. To simplify the problem we shall only deal with an astrophysically homogeneous group of stars, such as the K-giants, for many of which we have trigonometrically determined parallaxes; and the only assumption we shall have to make at present is that the mean absolute magnitude of the K giants is independent of their position in space.

The Allegheny parallaxes form the principal basis of our present parallax system, and we shall therefore discuss the 128 Allegheny parallaxes for K giants. In order to obtain absolute values, a correction of $+0''.003$ has been applied to the measured relative parallaxes. If the correction is too large or too small it should be found out in the course of this investigation.

If the Allegheny parallaxes are systematically too large the error will have a relatively larger influence on the smaller parallaxes. Accordingly, by reason of our first assumption, the mean absolute magnitude derived for the stars with small parallaxes may be expected to be fainter than that for the stars of large parallax.

Arranging the stars into three groups according to the size of their parallax, and deriving the mean reduced parallax (i.e., the mean value

of the parallaxes reduced to the apparent magnitude 5), we obtain the values given in table I.

TABLE I

LIMITS OF PARALLAX	NUMBER OF STARS	MEAN APP. MAG.	MEAN RED. PAR
$\pi > 0''.031$	26	4.03	$+0''.0315$
$0''.030 > \pi > 0.011$	56	4.29	$+0\ .0144$
$0\ .010 > \pi$	46	4.88	$-0\ .0005$

The apparent magnitude in all three groups being not greatly different, it is evident that the reduced parallax will "run" with the observed parallax. Therefore, we cannot, from table I alone, conclude anything as to the correctness of the trigonometric parallaxes because table I gives only a regression curve.

We can put our stars into groups according to apparent magnitude and again derive the mean reduced parallax. If the Allegheny parallaxes were systematically too large, and by a quantity independent of the size of the parallax, the effect would be greater for more distant stars and the numerical value of the reduced parallax should increase continuously with the numerical value of the apparent magnitude. This increase will be made more rapid on account of real dispersion in absolute magnitude. Table II gives the observed condition.

TABLE II

LIMITS OF APPARENT MAG.	NUMBER OF STARS	MEAN APP. MAG.	MEAN RED. PAR.
]3.0	11	2.61	$0''.006$
3.0 to 3.9	30	3.67	$0\ .014$
4.0 to 4.5	27	4.26	$0\ .011$
4.6 to 4.9	29	4.74	$0\ .014$
5.0 to 5.9	20	5.38	$0\ .017$
[6.0	11	6.54	$0\ .007$

We conclude that $\bar{\pi}(5.0)$, the mean reduced parallax, does *not* increase continuously with increasing m. The true relation between π and m must lie in between the two regression curves represented by tables I and II. In no case, therefore, can the reduced parallax *increase* with apparent magnitude. Accordingly, if the Allegheny parallaxes are at all affected by a general systematic error, we conclude from this that they are *not too large*, but possibly too small.

Corroboration of this is afforded by an analysis of the radial velocities and proper motions. For ninety-eight K stars we have radial velocities, proper motions, and Allegheny parallaxes. If the velocities of these stars are distributed at random it follows that

$$\bar{T}/\bar{V} = \pi/2, \quad \text{where } T = 4.74\ \mu/\pi.$$

If the Allegheny parallaxes were too large, we might expect the mean value of T to be smaller than $\pi/2\overline{V}$. As the Allegheny parallaxes reduced to absolute values still contain several negative and zero values, they cannot directly be used for the computation of mean T. The procedure arbitrarily followed, therefore, was to count the negative parallaxes as positive and to exclude the zero values entirely. This will tend to give too low a value for $\overline{T}$, and thus, if the mean T is still larger than $\pi/2\overline{V}$, it will strengthen our argument. In this way we find

$$\overline{V} = 21.4 \text{ km./sec}, \qquad \overline{T}_1 = 44.3 \text{ km./sec},$$

and $\overline{T}_2 = 33.7$ km./sec., if we also leave out the five largest values of T. If T is independent of the absolute magnitude, and therefore of the parallax, we have

$$\overline{T}_3\,\overline{\pi} = 4.74\overline{\mu}, \qquad \overline{T}_3 = 38.3 \text{ km./sec.}$$

All three values of T give a ratio $\overline{T}/\overline{V}$ in excess of $\pi/2$. Again we conclude that the parallaxes are probably not too large.

Once having adopted a "correct" system of parallaxes, we can proceed to determine the mean absolute magnitude of the K giants. To make our material still more homogeneous we shall take only the stars of Harvard spectral class K0. To increase the material, parallaxes determined at observatories other than Allegheny were also used, principally those determined at the McCormick Observatory. To the latter a systematic correction of $+0''.003$ was applied to reduce the observed to absolute values; likewise, $+0''.004$ was applied to the Allegheny parallaxes. In all, there are thus available 145 modern trigonometric values.

It is assumed that the distribution of absolute magnitudes is symmetrical with respect to the mean value and also that the same holds for the errors in the measured parallaxes. In both cases, then, the *mean* will coincide with the *median*, and, for stars of the same apparent magnitude, the *mean* absolute magnitude will correspond to the *median* parallax. When using stars of different apparent magnitudes the *mean* absolute magnitude will very nearly correspond to the *median* value of the reduced parallax.

The observations give:

$$\pi(5.0) \text{ median} = 0''.0136, \text{ and, accordingly, } \overline{M} = +0.67.$$

From the tables given in *Groningen Publications*, Nos. 29 and 34, we can compute the mean absolute magnitude for K0 stars of all apparent magnitudes and proper motions.

From Table 41 (*Gron. Publ.*, No. 29), we see that for apparent magnitudes 4.0, 5.0, and 6.0, we have $\overline{\pi} = 0''.025, 0''.018$, and $0''.011$, respectively (reduced to a speed of 21 km./sec. for the solar motion).

In view of the small probable error affecting these parallaxes, the correc-

tion from the absolute magnitude corresponding to mean parallax to a mean absolute magnitude will be small and not larger than $0^M.1$. Accordingly, we have for the mean absolute magnitude of the K0 stars with apparent magnitudes between 4 and 6, about $+1^M.1$. A small negative correction should be applied to this value to account for the inclusion in the Groningen material of a few K dwarfs among the stars brighter than the sixth magnitude. Probably, therefore, $+1^M.0$ is a more plausible value to adopt.

From the proper motions of the K0 stars in Boss's Catalogue, and from Table 37, *Gron. Publ.* No. 34, we find as the mean absolute magnitude for the K0 giants about $+0^M.6$.

Comparing the two values, +1.0 and +0.6, with that derived from trigonometric parallaxes, +0.7, we conclude again that these parallaxes cannot very well be considered too large.

The same treatment can be given to the M giants. As there are very few apparently bright M stars, and also since they are probably as a whole much brighter than the K giants, the measured parallaxes will be very small and systematic errors will have a large influence on the absolute magnitudes. We have at our disposal trigonometric parallaxes for forty-five giants, not including six long period variable stars. The median value of the reduced parallax, $\bar{\pi}(5.0)$, is $0''.0092$, corresponding to an absolute magnitude of $-0^M.2$. From the tables in *Gron. Publ.* No. 29, we find $-0^M.5$, and from those in *Gron. Publ.* No. 34, $-0^M.3$. Considering the scantiness of the parallax material and the unknown influence of selection upon it, we may consider the agreement as fairly good.

From the distribution of the reduced parallaxes, some conclusions may be drawn regarding the validity of the assumption that the distribution curves of the absolute magnitudes and measured parallaxes are symmetrical and also whether this distribution conforms to a first Laplacian error curve or to a second Laplacian or Gaussian error curve.

In figure 2 are plotted, on a logarithmic-arithmetic scale, the numbers of parallaxes larger than a certain value, beginning with the median. On this scale the first Laplacian curve, $e^{-|x|}$, is transformed into a straight line. We see that the observations lie on a straight line with remarkable closeness. For the parallaxes on the smaller side of the median the distribution does not accord as well with a Laplacian, as is shown in figure 1. In figure 3 the distribution of reduced parallaxes is shown on an arithmetic-probability scale, on which the Gaussian curve would be represented by a straight line. It appears from these figures that the distribution of the reduced parallaxes is *not* symmetrical, and that it shows a "beaky" maximum. In the present case, the distribution is a superposition of the distribution in absolute magnitude on that in the measured parallaxes, into which also enters the non-symmetrical distribution of apparent mag-

nitude. It thus appears likely that at least one of the first two distributions is not Gaussian.

Summarizing, we may say that the mean absolute magnitudes of the K0 and M giants are probably not far from $+0^M.7$ and $-0^M.2$, respectively. The agreement between these values and those derived from considerations of proper motion and parallactic motion strengthens the previous conclusion that the trigonometric parallaxes, at least those of Allegheny, are not affected by large systematic errors.

[1] van Rhijn, P., *Publ. Kapt. Astr. Lab., Groningen*, No. 34, 1923 (1–80).

MEASUREMENTS ON THE EXPRESSION OF EMOTION IN MUSIC

BY C. E. SEASHORE

STATE UNIVERSITY OF IOWA

Read before the Academy, November 15, 1922

All musical expression of emotion is conveyed in terms of pitch, intensity, duration, and extensity, which are the four elemental attributes of all sound. In their compound forms these attributes furnish us *timbre*, which on the basis of harmonic analysis is a complex of pitches and accounts for all the so-called qualities of sound from the pure tone to the purling of the brook; *rhythm*, which is a combination of intensity and duration effects in sound patterns, with all related time variants, such as tempo, staccato, and glide; *consonance*, which is the foundation for all phenomena of harmony and melody; and *volume*, which is the principal spatial and intensive aspect of sound.

This classification is probably complete, and we now have instruments and technique in the laboratory for the measurement of all these factors in terms of the sound wave; because the frequency of waves determines pitch, amplitude determines intensity, duration of the single wave determines extensity, and the form of the wave determines timbre. Thus, by combining measurements of frequency, amplitude, duration, and form of sound waves, we are enabled to take adequate account of every factor that enters into sound.

Now, when we stop to realize the situation, we find that everything in the way of musical expression that the singer conveys to the listener is conveyed in terms of the sound wave: when we eliminate sight and other senses which are merely accessory, there is only one avenue that can convey the musical message and that is the sound wave. The sound waves may be intercepted, recorded, measured, and analyzed by instruments

of precision so that we secure a detailed and faithful objective record of what the musician conveyed through this medium. Thus we can isolate, describe, and classify all types of variants, from the cold, non-emotional and mechanical production of tones to the most highly artistic expression of aesthetic emotion. This claim may sound crass and extravagant in the face of all our traditional regard for artistic expression of emotion as something elusive, indefinable, and intangible, and the general inability of experimental psychology to cope with the problem up to the present time But let us take some examples showing how it works out.

One of the most characteristic evidences of the tender emotion effectively expressed in singing is the vibrato. There is no end of confusion among musicians as to what the vibrato is, its desirability, how it may be acquired or eliminated, what it really means, and the factors that control it. It is, however, universally admitted that the vibrato represents an attempt to express emotion musically.

The study of the vibrato was first taken up by Dr. Schoen in our Iowa laboratory and is now continued by Mr. Kwalwasser. Recording and analyzing the sound waves to determine what constitutes the vibrato, we find that it is a synchronous pitch and intensity pulsation averaging about six oscillations per second. A particular vibrato can therefore be expressed adequately in terms of the three variables, pitch, intensity, and time; i.e., the range of pitch fluctuation, the range of intensity fluctuation, and the rate, regularity, and form of these two synchronous fluctuations. Within these three factors we shall then find all the possible variants of the emotional expression of the vibrato. Thus, for different singers, or for a given singer at different times, it may be primarily pitch pulsation; or primarily intensity pulsation; and regarding these as constant, the rate, the regularity, and the degree of approximation toward a sine curve in the pulsation are variables descriptive of emotional quality.

With all these factors under control, we may now take great singers and study the personal characteristic of each singer and the laws for the expression of different kinds of emotion through the vibrato. For this purpose phonograph records are of inestimable value because they produce the vibrato faithfully and thus furnish fair examples of the singing in vibrato when it was not known to the singer that this was a factor under observation. With the objective facts in hand, we only correlate the vibrato with principles of neural discharge showing the relation of artistic expression in music to nervous instability in terms of neurological concepts, for a tender emotion is a condition of nervous instability. We may here investigate the relation of a feigned emotion or a genuine emotion, according as the music was or was not actually expressed emotionally.

Another attribute of emotional expression is timbre. This is measured in terms of the form of the sound wave which may be analyzed into its

harmonic components by mechanical instruments and the harmonic elements may be combined into the complex wave form. Our question here resolves itself into the determination of types of the beautiful in timbre. These range, for various purposes, from the pure tone, as represented in the sine curve, to the richest tone reproducible, with significant peaks and valleys of affective tone. Types of beauty and ugliness may be traced and expressed in terms of the number and relative prominence of overtones.

Again, much of the emotional effect is expressed in terms of time, not only in rhythm and tempo and the various forms of acceleration or retardation, accentuation or holding of notes, but even more effectively in the form of attack and the release of tones.

These illustrations must suffice to show that everything in the nature of musical emotion that the musician conveys to the listener can be recorded, measured, repeated, and controlled for experimental purposes; and that thus we have at hand an approach which is extraordinarily promising for the scientific study of the expression of musical emotion.

It is, perhaps, superfluous to point out that in such study we are not concerned with the cause or character of the emotion of the musician except in so far as it is necessary for the understanding of what emotion is expressed. In the same manner, we are not concerned with the experience of emotion on the part of the listener except in so far as it is necessary for the understanding of what is actually transmitted in the tonal message.

The objective study of emotion has been approached profitably but remotely from the physiological point of view, as in the work on the glands of internal secretion and the autonomic system in general, as well as certain reactions through the skeletal muscles; but the bulk of our common knowledge of emotional expression rests upon crude observations of gross behavior in emotional situations, as in the study of animal behavior. The method of which I am speaking gives a new approach to a rigid and analytical behavioristic study of emotion in musical expression which will not only aid in the scientific analysis of musical emotion, but should throw valuable light on the fundamental laws of all emotional expression. There is nothing radically new in the technique; but, as in many another advance in science, the new step consists in realizing a new application of means already at hand.

Incidentally, it may be noted that this method may be used in retracing the emotional elements of primitive music as recorded by the phonograph, as many of the emotional features of musical expression are faithfully reproduced in such records. Heretofore we have had no objective method of evaluating this important element; a permanent graphic enlargement of the sound waves may be made from the fresh record, thus preventing deterioration and making the optical record available for detailed examination and measurement.

GROUPS OF ORDER 2^m IN WHICH THE NUMBER OF THE SUBGROUPS OF AT LEAST ONE ORDER IS OF THE FORM $1 + 4k$

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated July 28, 1923

1. *Abelian Groups.*—In this section we shall consider the general case when the abelian group G is of order p^m, p being an arbitrary prime number. It will be convenient to use the terms *higher type* and *lower type* as regards two distinct prime power abelian groups of the same order with the following meaning: If the invariants of these groups, arranged in descending order of magnitude, are $m_1, m_2, \ldots$, and $n_1, n_2, \ldots$, respectively, and if whenever the first difference between an m and the corresponding n occurs the value of m exceeds that of the corresponding n then the former group is said to be of a higher type than the latter. It may be noted that this distinction between higher and lower types is practically the same as that between higher and lower terms of a homogeneous symmetric function in n variables.

It will now be proved that whenever G involves subgroups of different types but of the same order then the number of the subgroups of the lowest type is always of the form $1 + p + kp^2$, where k is a positive integer or 0, whenever this number exceeds 1. In this case, the number of the subgroups of every higher type but of the same order is of the form kp^2. When there is only one subgroup of the lowest type for a given order then the number of the subgroups of next to the lowest type and of this order is of the form $p + kp^2$ while the number of the subgroups of every higher type and of this order is again of the form kp^2.

When there is only one subgroup of lowest type for a given order p^B this subgroup is characteristic under G, being composed of all its operators whose orders divide a given number. Each of the subgroups of order p^B which is of next to the lowest type in such a group must be contained in a subgroup of order p^{B+1} which involves the given subgroup of order p^B and of lowest type. The number of such subgroups contained in G is of the form $1 + p + kp^2$, whenever this number exceeds 1, and the number of the subgroups of order p^B and of next to the lowest type contained in such a subgroup of order p^{B+1} is of the form $p + kp^2$. As such a subgroup of order p^B appears in only one such subgroup of order p^{B+1} it has been proved that the number of the former subgroups is of the form $p + kp^2$ whenever there is only one subgroup of order p^B and of lowest type contained in G.

The fact that whenever G involves more than one subgroup of order p^B which is of lowest type then the number of such subgroups contained

in G must be of the form $1 + p + kp^2$ results from the fact that the operators of G whose orders divide the order of the largest operator contained in such a subgroup generate a characteristic subgroup of order $p^{B+\gamma}$, $\gamma > 0$, such that the p^{th} power of every operator of this subgroup is contained in the cross-cut of all the subgroups of order p^B and of lowest type. The quotient group of this characteristic subgroup as regards the cross-cut in question is of type (1,1,1,...), and the given subgroups of order p^B have a (1,1) correspondence with the subgroups of index p^γ in this quotient group. As the number of these subgroups is obviously of the form $1 + p + kp^2$ the theorem under consideration has been proved.

It remains to establish the fact that the number of the subgroups of order p^B which are of a type different from those considered above is always divisible by p^2. This results immediately from the fact that the number of ways in which the largest independent generators of such a subgroup can be selected from the operators of G is divisible by a power of p whose index is at least two units higher than the index of the highest power of p which divides the number of ways in which such a generator can be selected from the operators of such a subgroup. The number of ways in which the other independent generators (if any) can be selected from the operators of G is evidently divisible by at least as high a power of p as the number of ways in which such generators can be chosen from the operators of the subgroup in question. Hence the general theorem announced in the second paragraph has been completely established. In particular, it may be noted that *the number of the subgroups of every order greater than 1 contained in a non-cyclic abelian group of order p^m is always of the form $1 + p + kp^2$, p being any prime number.*

From the method of proof employed above it follows directly that if the highest invariant of a subgroup of order p^B is p^k times the largest invariant of a subgroup of lowest type and of this order then the number of the subgroups which are of the same type as the former is always divisible by p^k. When G involves more than one subgroup of lowest type the number of the subgroups of the former type is always divisible by p^{k+1}. This is evidently an extension of the theorem noted in the second paragraph.

2. *Non-Abelian Groups.*—From the preceding section it results that there is no non-cyclic abelian group of order 2^m which involves subgroups of an order greater than 1 whose total number is of the form $1 + 4k$, and it has recently been proved[1] that there is no non-cyclic group of order p^m, p being an odd prime number, which involves subgroups of a given order greater than 1 whose total number is of the form $1 + kp^2$. Hence the only non-cyclic prime power groups which involve subgroups of at least one order greater than 1 whose total number is of the form $1 + kp^2$ must be non-abelian groups of order 2^m. Moreover, the subgroups of

index 2 involved in such a group must always be of the form $3 + 4k$. On the other hand, in the dihedral groups, whose order is of the form 2^m there is evidently no subgroups of any other order whose total number is of this form. That is, in such a dihedral group the number of subgroups of every index greater than 2 is of the form $1 + 4k$.

If a non-cyclic group of order 2^m contains an even number of cyclic subgroups of every order which divides the largest invariant of this group and exceeds 2 then the number of its subgroups of every order greater than 1 is of the form $3 + 4k$. This follows almost immediately from the theorem that every non-cyclic group of order p^m, p being any prime number, which contains an invariant subgroup giving rise to a cyclic quotient group of order p^α, $\alpha > 1$, must involve a multiple of p such invariant subgroups. In fact, the invariant subgroup in question involves the commutator subgroup but it is not coincident with it since the commutator quotient group must be non-cyclic. It must therefore correspond to a subgroup of index p^α in this quotient group which involves the $p^{\alpha\text{th}}$ power of every operator of this quotient group. The subgroups of this quotient group which give rise to non-cyclic quotient groups of order p^α have for their cross-cut the $p^{\alpha-1}$ power of every operator in the commutator quotient. As the index of this cross-cut under the group cannot be less than p^α the number of the subgroups which give rise to such non-cyclic quotient groups is of the form $1 + kp$. Hence the number of those which give rise to cyclic quotient groups of order p^α must be a multiple of p.

From the preceding paragraph it results that the only non-cyclic groups of order 2^m in which the number of the subgroups of at least one order greater than 1 is of the form $1 + 4k$ are contained among those which involve an odd number of cyclic subgroups of composite order. It is known that there are three and only three such groups for every value of $m > 3$, and when $m = 3$ there are two such groups.[2] Each of these groups contains subgroups of at least one order whose number is of the form $1 + 4k$, k being a positive integer or 0. In fact, besides the identity and the subgroups of index 2 the number of the subgroups of every order is of this form. Hence the theorem: *There are three and only three groups among the total number of the possible non-cyclic groups of order p^m, p being any prime number and $m > 3$, in which the number of subgroups of at least one order is of the form $1 + kp^2$. These three groups are characterized by the facts that they are non-cyclic, of order 2^m, and involve an odd number of cyclic subgroups of composite order.*

[1] G. A. Miller, *Proc. Nat. Acad. Sci.*, **9**, 237 (1923).

[2] G. A. Miller, *Trans. Amer. Math. Soc.*, **6**, 55 (1905).

THE MOTION OF A FALLING CHRONOGRAPH PROJECTILE

BY L. THOMPSON*

DEPARTMENT OF PHYSICS, KALAMAZOO COLLEGE

Communicated, July 9, 1923

Several chronographs depend upon the motion of a falling projectile, notably the Le Boulengé and the Drop Chronograph.[1] The application of the former is mainly in the measurement of relatively long time intervals, of the order of several hundredths of a second, since the means of producing a record does not permit sufficient precision for the determination of small differences. The Drop Chronograph gives an accurate value

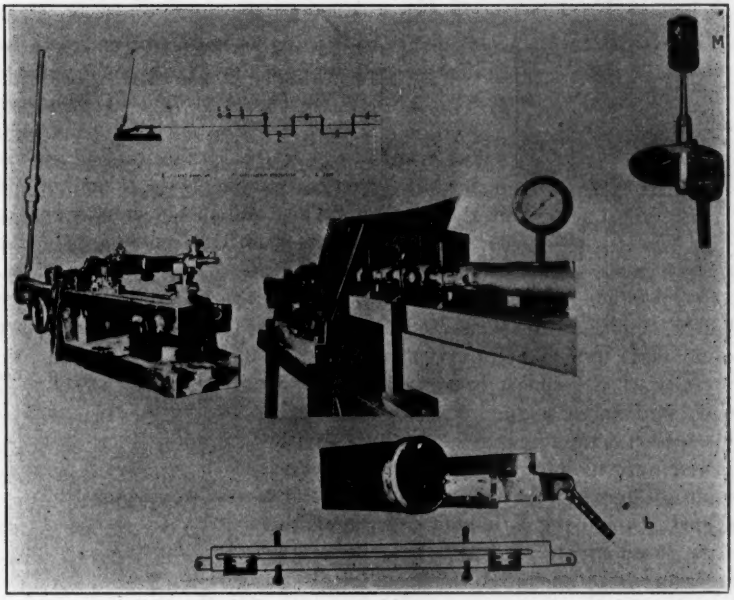


FIGURE 1

for very short intervals, for example a hundred-thousandth of a second in terms of the heights of fall to two sets of levers which control electrical recording circuits.

The apparatus shown in figure 1 is an optical chronograph, constructed for certain exterior ballistic experiments.[2] Neither the high-velocity projectile nor that of the chronograph, the latter carrying the photographic surface upon which the record is made, strikes a material object during the

experiment. When the pendulum is released, the circuit of the magnet M is broken and the chronograph projectile begins to fall. The gun is discharged an instant later by impact of the pendulum with the firing lever. The flight of the rifle projectile can be well timed in this way, with reference to the motion of the chronograph, the interruptions in the trace on the photographic surface recurring at practically the same place with successive firings. The instrument in figure 1*b* is employed to measure the distance between adjacent crossings of the light through the trajectory, this being approximately 70 cms. Small photographic plates are held securely against the straight edges of the two frames one mounted at each end of the carriage. The carriage has supporting wheels on which it can be drawn through the large brass tube, and is moved to a point where the image of the slit source is formed on one of the plates. Half of the light passes over the top of this surface and forms the image at the next crossing on the plate at that end of the carriage. The total distance between the crossings is found from the separation of the straight edges, and the distance from the lines to the ends of the plates.

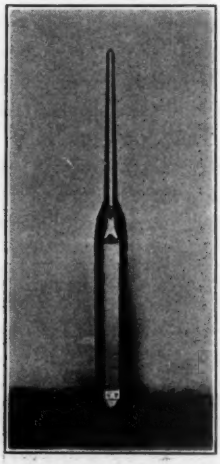


FIGURE 1*b*

In order that the change in velocity in a short section of the trajectory shall be significant for the determination of the air resistance, it is necessary to eliminate, or make correction for, small disturbing factors ordinarily regarded as negligible in chronographic work. Two corrections of this nature, heretofore neglected, which are considered in the use of this chronograph, are effects resulting from (*a*) the influence of the holding magnet M upon the early stages of the motion of the chronograph projectile, the acceleration increasing from an initial value of zero to the final value g during the breaking of the magnet circuit; and (*b*) the resistance of the air in slightly retarding the falling projectile. Both of these factors determine corrections which are larger than the accidental errors in the operation of the chronograph, which is capable of registering with an error otherwise not greater than 1/2,000,000th of one second. A drum camera of the usual type for film records was abandoned because of the irregular change in length of the film which occurs

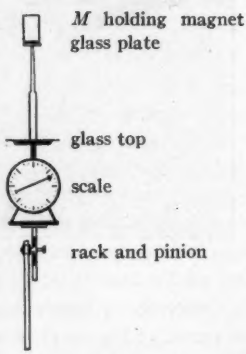


FIGURE 2

during development and drying. This is likely to prevent realization of the required precision, even though such factors as sharpness of interruption and reliability of time-calibration instruments are not questioned. Aluminium projectiles are used to increase the velocity decrement, and the tube provides a means of compressing the air or other gases for experiments at higher densities.

To determine the effect of the magnet, the chronograph projectile was weighed at various points near M with the apparatus shown in figure 2. For points very near M the system is unstable. At these positions plane glass plates were placed between the top of the projectile and M, the rack being lowered until the projectile was about to drop away. In this condition of equilibrium the effective weight is read from the scale, correcting for the weight of the glass plate. The result of this experiment is given in figure 3. Curve I was made from data obtained while using a current in M just sufficient to hold the projectile when in direct contact, that is, its value at the instant of release. Curve II was drawn from an oscillogram of the current in M during the release, by changing the horizontal coördinate from t to h, the distance of fall. Curve III is the result of multiplying the ordinates of I and II, and gives the value of the variable acceleration $G(h)$ by measuring from the upper horizontal line.

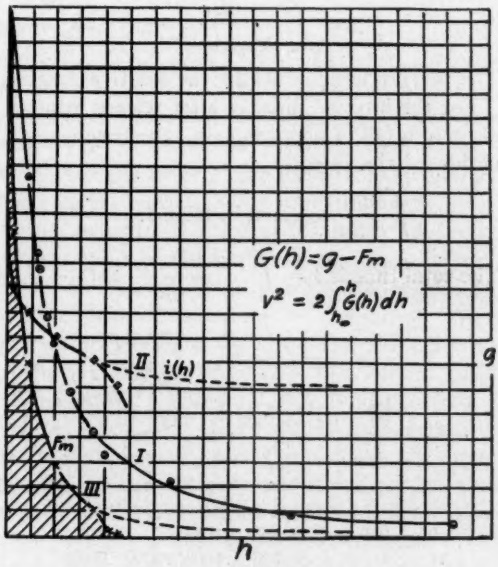


FIGURE 3

The equation of motion for the falling projectile is

$$\frac{dv}{dt} = g - \frac{\mu}{(h+h_0)^2} = g\left(1 - \left(\frac{h_0}{h+h_0}\right)^2\right); \text{ hence } v = \sqrt{2gh}\cdot\sqrt{\frac{h}{h+h_0}},$$

and the interval of time utilized on the chronograph record is then

$$\Delta t = \frac{1}{\sqrt{2g}}\left\{2\left(\sqrt{h+\Delta h}-\sqrt{h}\right) - h_0\left(\frac{1}{\sqrt{h+\Delta h}}-\frac{1}{\sqrt{h}}\right)\cdots\right\} =$$

$$\Delta t_a\left(1+\frac{h_0}{\sqrt{h_1 h_2}}+\ldots.\right).$$

The interval Δt_a is the uncorrected value of the time for a body falling freely through a distance Δh from a height h. To represent the data of Curve III according to the equation of motion, h_0 is approximately 0.1 cm.

The following statement suggests a more satisfactory means of computation. The shaded area under Curve III is evidently proportional to the unacquired kinetic energy by virtue of the effect of M. Hence

$$v^2 = 2gh - 2\int F(h)dh \text{ and, for } h > 0.5 \text{ cm.}, \quad 2\int F dh = \overline{K} = 155.\left(\frac{\text{cm.}}{\text{sec.}}\right)^2,$$

$$\Delta t = \int_{h_1}^{h_2} \frac{dh}{\sqrt{2gh - \overline{K}}} = \Delta t_a(1 + \overline{K}/4g\sqrt{h_1 h_2} + \ldots).$$

The second term indicates the relative magnitude of the correction.

For the interval directly after release, while M is still effective, equivalent to a fall of about 0.5 cm., the time is obtained by a graphical evaluation of the integral

$$t = \int \frac{dh}{v(h)} \text{ as given in table I.}$$

The total time is $t = \int_0^{h=.5} \frac{dh}{v(h)} + \Delta t_1(1 + \overline{K}/4g\sqrt{.5h} + \ldots),$

TABLE I

h cm.	$v = \sqrt{2gh}$	$v = \sqrt{2(gh - \int F dh)}$	$\frac{1}{v}$	$\Delta t_1 = \int_{.05}^{h} \frac{dh}{v(h)}$	$\Delta t_c = \frac{1}{g}(\sqrt{2gh})_{.05}^{h}$	$\Delta t_a = \frac{(\sqrt{2gh - \overline{K}})_{.1}^{h}}{g}$	$\Delta t_1 + \Delta t_2$	$\Delta t - \Delta t_c$	$\frac{\delta \Delta t}{\Delta t}$
.05	9.9	5.902	.1694						
.1	14	9.916	.1008	.0060	.0042			.0018	
.2	19.8	15.986	.0625	.0137	.0101			.0036	
.3	24.25	20.763	.0482	.0192	.0146			.0046	
.4	28	24.862	.0402	.0236	.0185			.0051	
10	140	139.4453			.1328	.1164	.1400	.0072	.051
20	197.9901	197.5981			.1919	.1757	.1993	.0074	.037

					$\Delta t_c = \frac{1}{g}(\sqrt{2gh})_{h_1}^{h_2}$	$\Delta t = \frac{(\sqrt{2gh - \overline{K}})_{h_1}^{h_2}}{g}$			$K/4g\sqrt{h_1 h_2}$
10									
20			Le Boulengé		.0592	.0594		.0002	.0028
150	542.2177	542.0747							
155	551.1805	551.0399	Optical Chr.		.009145	.009148		.000003	.00026

TABLE II

INTERVAL	h_2-h_1	ε^2	h_3-h_2	ε^2	h_4-h_3	ε^2
a	12.8001	100	12.7980	81	12.6557	784
b	12.7980	121	12.7976	25	12.6616	961
c	12.8002	121	12.7962	81	12.6573	144
d	12.7991	000	12.7987	256	12.6588	9
e	12.7982	81	12.7950	441	12.6592	49
Mean	12.7991		12.7971		12.6585	
$\sqrt{\frac{\Sigma\varepsilon^2}{n}}$		.0009		.0013		.0019
Velocity equivalent for velocities of order of 2500 ft./sec.		±.2		±.3		±.4

where Δt_1 is the uncorrected time of fall from a point where M is no longer effective.

Table I contains numerical values for ranges characteristic of the chronographs as generally used. The correction for the total time of fall is large, as would be expected from the effect of the magnet upon the early motion. This is well shown by the corresponding velocities in the second and third columns, the former being for the case of free fall. The corrections for the differences in the times of fall are relatively small. After a fall of 150 cms., the value for an interval of 0.01 second is approximately 0.000003 second. For the air-resistance work this is significant. With a typical Le Boulengé interval, for example, ten to twenty cms. below the disjunctor reading, the correction is of the order of 0.0002 second (using the K for the optical chronograph). The actual Le Boulengé error is likely to be larger, for other reasons. Also, the two holding magnets of the instrument may have a relative value of K during the disjunctor experiment which differs from that for the ordinary use of the chronograph. This is true even if the same initial currents are involved, and the effect is amplified if the two holding systems release at different currents.

Since the chronograph velocities are very low, the effect of air resistance can be included with good approximation by writing

$$\frac{dv}{dt} = f(v) = g - cv$$

the integral of which is approximately

$$\Delta t = \Delta t_a\left(1+\frac{c}{6}\,(t_{a_2}+t_{a_1})+\ldots\right), \quad \text{where} \quad t_a = \sqrt{\frac{2h}{g}}.$$

The constant c was determined experimentally by placing a wood copy of the chronograph in a small wind tunnel, the force being measured by a roof balance. The air velocities were taken with a pitot tube. The value found for c was 0.003 (1/sec.). The second term, representing the correction, is about 1/2000th of the first term.

Both corrections will primarily affect the absolute values rather than the "law" of retardation.

As examples of the type of record obtained with the optical apparatus, enlargements of two of the older films are reproduced in figure 4, with the measurements of one of them, 4*a*, arranged in table II to indicate the precision in locating the base lines of the interruptions. This is the limiting factor in determining the possible precision. Figure 4*b* is included to show the appearance of the record in case of tumbling or unstable flight. It may be noted that the axis has changed from a position of large deflection in the horizontal plane at crossing 1 to a nearly normal horizontal position at 3, having a vertical component at 2. This would correspond to at least one half-cycle if the motion were an ordinary precession. This was taken with a 0.3 inch aluminium projectile. Each set of values of the differences in the rows *a*...*e* of the table is the mean obtained from six series of readings of the crossings 1 to 4 in order. This procedure reduces the tendency to recall the previous positions of the cross hair with respect to the base lines, and provides a fairly independent series of estimates of the line of separation. The sharpness of this line depends upon the careful focusing of the last lens of the set as the light is brought to the chronograph projectile, and the use of a narrow slit source with sufficient light. The last lens is mounted on a screw so that it can be focused with a micrometer adjustment.

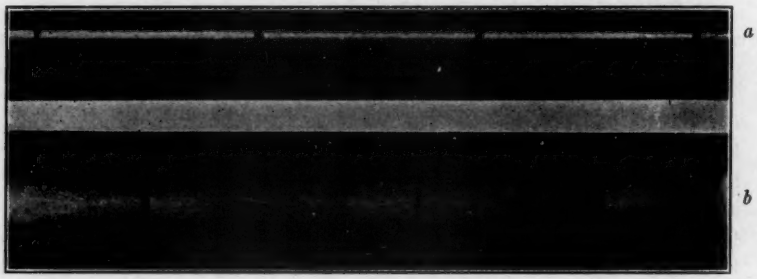


FIGURE 4

* NATIONAL RESEARCH FELLOW.

[1] Webster, A. G., "An Experimental Determination of the Period of Electrical Oscillations," *Physic. Rev.*, **6**, 1898; Hubbard, J. C., *Ibid.*, (2) **1**, 1913.

[2] "The Resistance of a Compressible Fluid at High Velocities," *Ibid.*, **17**, 1921.